

Using Convolutional Codes for Key Extraction in Physical Unclonable Functions

Sven Muelich and Martin Bossert

Institute of Communications Engineering, Ulm University, Germany
sven.mueelich@uni-ulm.de, martin.bossert@uni-ulm.de

Abstract. Physical Unclonable Functions (PUFs) exploit variations in the manufacturing process to derive bit sequences from integrated circuits which can be used as secure cryptographic keys. Instead of storing the keys, they can be reproduced when needed. Since the reproduced sequences are not stable, error correction must be applied. Recently, convolutional codes were used for key reproduction. This work shows that using soft information at the input of the decoder and list decoding reduces the key error probability compared to results from the literature.

1 Motivation

Physical Unclonable Functions (PUFs) can be used to generate and reproduce secure cryptographic keys. Since key reproduction is erroneous, methods for error correction are applied. In [1], convolutional codes were used the first time for PUFs. [2] uses a modified Viterbi decoder for convolutional codes, which outputs not only the decoding result, but also uses *Reliable Output Viterbi Algorithm* (ROVA) to produce reliability information about the result at the output of the decoder. Also, [2] showed how to increase efficiency by using multiple readouts. If the decoded sequence is too unreliable, a second and third readout are performed and the most reliable one is chosen. Using this approach, the authors were able to reduce both, the number of PUF response bits and helper data bits which have to be used in order to reproduce a key with a certain error probability. In this work, we apply further methods from the field of convolutional codes to the key regeneration process and compare the results to the approach chosen in [2].

Section 2 introduces the concept of PUFs and states the problem considered in this work. Section 3 illustrates the fundamentals of convolutional codes. In Section 4, the coding theoretical concepts which are applied to the PUF scenario for the first time in this work are explained. Section 5 presents our results and compares them to results from the literature. Section 6 concludes the paper.

2 Physical Unclonable Functions (PUFs)

Generating cryptographic keys by using *Pseudo Random Number Generators (PRNGs)* and storing them in non-volatile memory has several drawbacks. First, PRNGs do not provide true randomness and hence violate the demand for unique

and unpredictable keys. Second, storing keys in memory makes them vulnerable to physical attacks and hence requires expensive protecting mechanisms which need additional chip area and power consumption. PUFs are devices which extract unique cryptographic keys from intrinsic randomness like delay characteristics of digital circuits [3] or the initialization behavior of memory cells [4]. All devices (even with the same functionality) differ regarding this intrinsic randomness, which is unique for each device and cannot be controlled during the manufacturing process due to technical and physical limitations. A PUF uses this randomness to extract a random, unique and unpredictable sequence of bits which can be used as cryptographic key. Since the randomness is reproducible, the key can simply be regenerated when needed instead of storing it permanently. Hence, PUFs implement secure key generation and secure key storage.

2.1 SRAM PUF

SRAM PUFs were introduced in [4]. Dependent on the mismatch of two inverters of a SRAM cell (which is caused by manufacturing variations) and electrical noise, every SRAM cell converges to either 0 or 1 when powering-up the device. Since the mismatch is static and dominates over the noise, most cells always power-up with either 0 or 1. However, some of the cells do not have a preferred behavior and hence sometimes initialize with 0 and sometimes with 1. It is possible to produce reliability information of the individual cells by obtaining statistics during an enrollment phase. Based on such statistics, [5] derived a cumulative density function (cdf) and a probability density function (pdf) which approximate the measured one-probability P_{one} of SRAM cells:

$$\mathbf{cdf}_{P_{one}}(x) = \Phi(\lambda_1 \cdot \Phi^{-1}(x) - \lambda_2) \quad (1)$$

$$\mathbf{pdf}_{P_{one}}(x) = \frac{\lambda_1 \cdot \varphi(\lambda_2 - \lambda_1 \cdot \Phi^{-1}(x))}{\varphi(\Phi^{-1}(x))} \quad (2)$$

In Equations (1)–(2), Φ denotes the standard normal cumulative distribution function. For these functions we use $\lambda_1 = 0.51$ and $\lambda_2 = 0$ according to [5], in order to obtain a model which is equivalent to a *binary symmetric channel (BSC)* with crossover probability $p = 0.15$, which is often used in the PUF scenario. [5] also derived a cumulative density function and a probability density function¹ of the error probability P_e of an SRAM cell:

$$\mathbf{cdf}_{P_e}(x) = \mathbf{cdf}_{P_{one}}(x) + 1 - \mathbf{cdf}_{P_{one}}(1 - x) \quad (3)$$

$$\mathbf{pdf}_{P_e}(x) = \begin{cases} \mathbf{pdf}_{P_{one}}(x) + \mathbf{pdf}_{P_{one}}(1 - x) & , \text{ if } x < 0.5 \\ 0 & , \text{ otherwise} \end{cases} \quad (4)$$

Figure 1 visualizes both probability density functions, $\mathbf{pdf}_{P_{one}}$ and \mathbf{pdf}_{P_e} .

¹ Note that the pdf of the error probability is incomplete in [5], since there it does not include the case differentiation (cf. Equation 4) which is necessary.

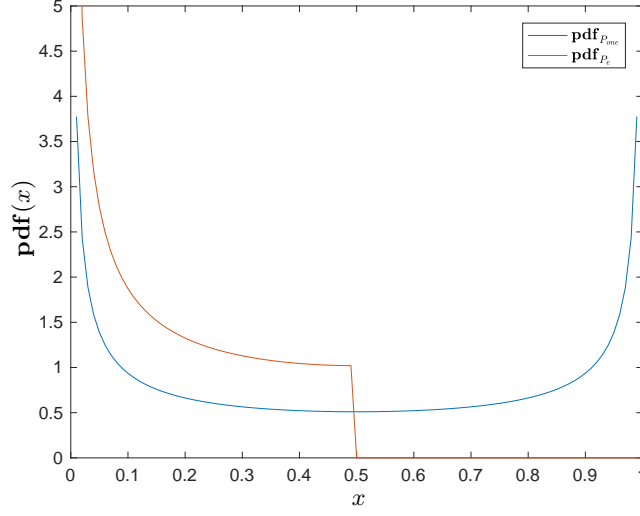


Fig. 1. Density functions of one-probability and error probability ($\lambda_1 = 0.51$, $\lambda_2 = 0$).

2.2 Helper Data Generation

Unfortunately, the derived bit sequences suffer from environmental conditions like temperature or voltage supply. These parameters cause, that responses may vary slightly. To circumvent this problem, *error correcting codes* can be used. Therefore, the initial PUF response is used to obtain *helper data*, which are needed for key reproduction. Since the helper data do not reveal information about the key, they do not need a protected memory. Extraction of helper data and using them to reproduce a key can be performed by different *helper data schemes*, cf. [6, 7, 1, 8].

2.3 Problem Statement

We start our studies based on the scenario in [2]: We extract t response bits from the PUF. Using the model introduced in Section 2.1, each of these t bits can be modeled as a BSC with crossover probability p_i ($i = 1, \dots, t$). We select $s \leq t$ Bits r_i ($i = 1, \dots, s$) which belong to BSCs with crossover probabilities $p_i < p_T$, where p_T is a chosen threshold. Thus, unreliable PUF bits are ignored. [2] uses convolutional codes and ROVA in order to extract and reproduce keys. The reliability information of the several bits is only used in order to identify reliable bits within the described scenario. Our approach additionally translates these information into soft information which are used at the input of Viterbi to improve the decoding results by decreasing the key error probability P_{err} . Additionally, we use a list decoding variant of Viterbi for further improvements. For comparability with [2], we also use codes of rate $\frac{1}{2}$.

In a second step we use all extracted bits without refusing the unreliable ones. Our goal is to study, whether the good channels compensate the bad ones when using soft information at the input of the decoder. For a fair comparison we have to adjust the code rate for this scenario.

3 Convolutional Codes

Convolutional codes, introduced by Elias [9], are widely used in applications. Efficient methods for implementations in hardware make them suited for PUFs. [1] applied convolutional codes to PUFs for the first time. An optimized FPGA Viterbi decoder adapted to PUFs was designed in [10]. We explain encoding and decoding briefly, for a detailed description of convolutional codes we refer to [11, Chapter 8].

3.1 Encoding

Convolutional codes can be described using the parameters $(n, k, [\mu])$. Using a code of rate $\frac{k}{n}$, we partition an information sequence into blocks of length k . Each information block is mapped to a codeword of length $n > k$. In contrast to block codes, codeword \mathbf{c}_r depends on information block \mathbf{i}_r and on the μ previous information blocks $\mathbf{i}_{-1}, \dots, \mathbf{i}_{-\mu}$.

Encoding of an information block can be done by calculating $\mathbf{c}_r = \mathbf{i}_r \mathbf{G}_0 + \mathbf{i}_{r-1} \mathbf{G}_1 + \dots + \mathbf{i}_{r-\mu} \mathbf{G}_\mu$, or equivalently $\mathbf{c}_r = (\mathbf{i}_r, \mathbf{i}_{r-1}, \dots, \mathbf{i}_{r-\mu}) \cdot (\mathbf{G}_0, \mathbf{G}_1, \dots, \mathbf{G}_\mu)^T$, using $(k \times n)$ generator matrices $\mathbf{G}_0, \mathbf{G}_1, \dots, \mathbf{G}_\mu$. This process can be implemented by using a linear shift register with k inputs, n outputs and μ memory elements, cf. Figure 2 for a $(2, 1, [2])$ encoder with $\mathbf{G}_0 = (1, 1)$, $\mathbf{G}_1 = (1, 0)$, $\mathbf{G}_2 = (1, 1)$. Initially, all memory elements contain zeros. Information block \mathbf{i}_r is mapped to codeword $\mathbf{c}_{r,1} \mathbf{c}_{r,2}$ by applying linear operations. Then all bits move one memory element to the right. Bit \mathbf{i}_r enters the leftmost memory element, bit \mathbf{i}_{r-2} is dropped. Encoding continues with block \mathbf{i}_{r+1} .

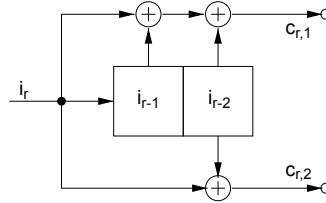


Fig. 2. Encoder of a $(2, 1, [2])$ convolutional code.

3.2 Decoding

The Viterbi algorithm realizes efficient *maximum-likelihood decoding* [12]. We can visualize decoding using a graph structure called *trellis* (cf. Figure 3 for the trellis of the $(2, 1, [2])$ code introduced in Section 3.1). The nodes in the trellis represent the 2^μ states of the shift register. Each node has two outgoing edges for encoding a 0 (dashed lines) or a 1 (solid lines). Since the register is initialized with zeros, not all states occur in the beginning. The edges are labeled with the corresponding codeword and represent the transitions to the next memory state. Hence, each path in the trellis represents a possible code sequence \mathbf{c} . The decoder aims to find the code sequence which was transmitted with highest probability when the sequence \mathbf{y} was received, hence it maximizes $p(\mathbf{y}|\mathbf{c})$.

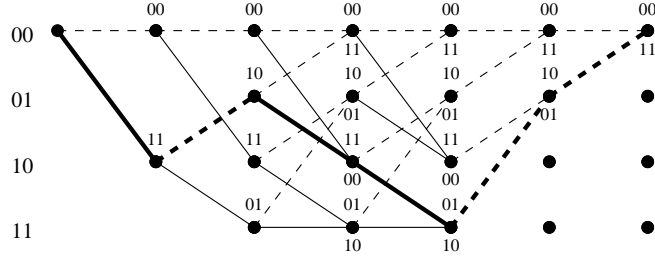


Fig. 3. Trellis for decoding the $(2, 1, [2])$ code from Section 3.1.

Let $\mathbf{c} = (c_1, c_2, \dots)$ denote the code sequence which is sent over a noisy channel, \mathbf{c}_i denotes a length- n block. Analog to this notation, \mathbf{y} denotes the received sequence. We define a *submetric* and a *metric*.

Definition 1. For every edge, the submetric is defined as

$$\lambda_i = n - \text{dist}_H(\mathbf{c}_i, \mathbf{y}_i),$$

where $\text{dist}_H(\mathbf{c}_i, \mathbf{y}_i)$ denotes the Hamming distance of \mathbf{c}_i and \mathbf{y}_i . The metric of a node is the sum over all η submetrics λ_i on a path, i.e. $\Lambda = \sum_{i=1}^{\eta} \lambda_i$.

Viterbi calculates the metric for all nodes in the trellis. To calculate the metric for a specific node, for each incoming edge Viterbi adds the incoming submetric to the metric of the corresponding node on the left side of the segment. Viterbi stores the maximum reached metric for the node as its metric and adds the edge which contributes to this metric to its survivor path. To circumvent error propagation, often m tail bits are inserted after every L information bits to reset the register (cf. Figure 3, trellis segments 5 and 6). The bold path in Figure 3 is the survivor path, which gives us \mathbf{c} with $p(\mathbf{y}|\mathbf{c})$ maximized.

4 Techniques used in our Approach

We focus on SRAM PUFs, since it is easy to derive reliability information for this type as shown in [13]. For our studies we use the model stated in Section 2.1, Equations (1)–(4).

4.1 Using List Decoding

The simplified ROVA, introduced in [14] and used in [2], has the drawback, that it calculates error probabilities for each symbol in order to compute the total error probability for the whole code sequence. As alternative to the calculation of error probabilities, we suggest to use a list decoding variant of the Viterbi algorithm according to [15], which determines the \mathcal{L} most reliable paths in the trellis.

Using this modification of the Viterbi algorithm, each node stores the difference between the metric of the survivor and the second best path in addition to the metric. When the maximum likelihood path was determined, the minimum of the metric differences of all nodes on the ML path is selected, since the most unreliable decision was taken at the corresponding node. At that node, we take the non-surviving edge to differ from the surviving path, and follow the second path backwards until it merges again with the survivor path. By recursively applying this decision process, we can generate the \mathcal{L} most likely code sequences.

4.2 Using Soft Input Information for Viterbi

Instead of using the SRAM reliability information (cf. Section 2.1) only for choosing reliable bits from the PUF response, we can additionally translate them into soft information which can be used at the input of Viterbi to improve the decoding results. We calculate soft information

$$s_i^{(j)} = (-1)^{r_i^{(j)}} \cdot (\log(1 - p_{e_i}) - \log(p_{e_i})) \quad (5)$$

for cell i at time instance j , where r_i is the i -th response bit and p_{e_i} is the error probability of SRAM cell i . This formula is already given in [5], however the authors use short linear block codes with soft decision decoding instead of convolutional codes. We obtain p_{e_i} for all cells i by using the model described in Section 2.1. In a practical scenario the soft information can be obtained during the enrollment phase by doing measurements for every individual PUF.

5 Experiments and Results

We perform simulations to compare our ideas to the results in [2]. According to [2], we aim for an error probability in the order of 10^{-6} .

5.1 Soft Input and List Decoding instead of ROVA

In a first experiment, we use the same scenario as in [2], using thresholds $p_T \in \{0.3, 0.2, 0.1\}$. However, instead of ROVA we use list Viterbi with list size \mathcal{L} (cf. Section 4.1). Also, we translate information about the several channels into soft information (cf. Section 4.2), which we use at the input of Viterbi according to Equation (5). For comparability with [2], we use the $(2, 1, [6])$ and $(2, 1, [7])$ convolutional codes. Additionally, we use rate $\frac{1}{2}$ convolutional codes with memory length 10, 14, and 16 to further improve the results.

The results in Table 1 show, that using soft information as input results in smaller error probabilities compared to ROVA. Even without list decoding, the use of soft information at the input of Viterbi yields an improvement (cf. column “SD, $\mathcal{L} = 1$ ”). Applying list decoding further decreases the error probability (cf. columns “SD, $\mathcal{L} = 3$ ” and “SD, $\mathcal{L} = 4$ ”). Convolutional codes with larger memory length reduce the error probability essentially, but at the cost of a larger runtime. However, runtime is often secondary in PUF applications. Columns “HD, $\mathcal{L} = 3$ ” and “HD, $\mathcal{L} = 4$ ” show, that only using list decoding without soft information at the input is not sufficient to improve the results in [2]. Note that we obtain the reference error probabilities by taking the mean over the corresponding bins of the binning approach used in [2].

Figure 4(a) compares the key error probability of the $(2, 1, [6])$ code from [2] with our soft information input and list decoding approach for different values of p_T . Note that the threshold measure used in [2], p_{max} , is specific for the *Differential Sequence Coding (DSC)* scheme implemented in that reference, and hence is translated to p_T in Tables 1 and 2.

5.2 Additional Use of Multiple Readouts

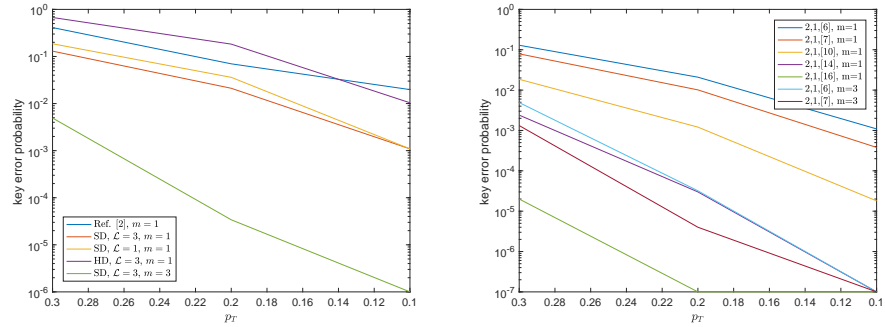
The goal in this section is to further decrease the error probabilities gained in Section 5.1, by additionally using multiple readouts as done in [2]. Table 2 presents the error probabilities obtained from simulations with list size \mathcal{L} and using m readouts. Applying multiple readouts results in an essential improvement. We obtain a lower error probability at the cost of extracting more bits and multiple decoding processes. A $(2, 1, [6])$ code with $p_T = 0.1$ results in an error probability of 10^{-6} used with list size $\mathcal{L} = 3$ and $m = 2$ readouts (or even without list decoding and $m = 3$ readouts). For convolutional codes with larger memory size, p_T can be larger in order to achieve $P_{err} \approx 10^{-6}$. Figure 4(b) compares convolutional codes with different memory length, using soft information input, list size 3 and multiple readouts. Compared to [2], our approach also results in error probabilities smaller than 10^{-6} for $p_T = 0.1$.

5.3 Without Refusing Unreliable Bits

A drawback of the approaches described so far is, that not all response bits are used. Hence more bits than needed have to be extracted from the PUF, what can be time-consuming. Instead of refusing response bits with an error

probability higher than a certain threshold p_T , we can use all response bits and use their reliability information as soft input to a Viterbi decoder. The idea is, that the soft information of reliable bits compensate unreliable bits. For a fair comparison with [2], we have to adjust the code rate to $\frac{1}{3}$, since in contrast to the experiments performed in Sections 5.1 and 5.2 we use all extracted bits now.

The results in Table 3 show that this method significantly improves the previous approaches. For the implementation which is specifically chosen in [2], this means two essential improvements: First, the helper data in the DSC scheme can be further reduced by factor $\frac{1}{2}$ since one vector is only used for identifying reliable bits from the response. Second, ROVA and binning can be avoided.



(a) (2,1,[6]) code with soft information at (b) Convolutional Codes with different input, list decoding, and m readouts. memory lengths and m readouts.

Fig. 4. Comparison of convolutional codes with different memory lengths, soft input information, list decoding with list size \mathcal{L} , and m readouts.

6 Conclusion

This work studied how previous approaches that use convolutional codes for PUFs can be improved by using soft information input for Viterbi and list decoding. Our results show, that the selection of reliable bits from a PUF response is not needed. Reliability values of the individual response bits are translated into soft information which is used at the input of the Viterbi. Using these soft information, reliable channels compensate unreliable ones. A rate $\frac{1}{3}$ convolutional code of memory length 7 is sufficient in order to result in a key error probability of order 10^{-6} . The obtained error probabilities can further be decreased by increasing memory length or using a larger list size for list decoding. Using a convolutional code with more than 8 memory elements, the error probability becomes smaller than 10^{-7} . To summarize, our strategy has two advantages compared to previous approaches. First, all response bits are used. Second, we avoid overhead since our work does not require ROVA or any binning approach.

Code	pr	Extractions	Ref.[2]	HD, $\mathcal{L} = 3$	HD, $\mathcal{L} = 4$	SD, $\mathcal{L} = 1$	SD, $\mathcal{L} = 3$	SD, $\mathcal{L} = 4$
(2,1,[6])	0.3	500.000	4.11e-01	6.73e-01	6.67e-01	1.86e-01	1.30e-01	1.2701e-01
		10.000.000	–	6.72e-01	6.66e-01	1.86e-01	1.29e-01	1.27e-01
	0.2	500.000	6.98e-02	1.83e-01	1.70e-01	3.61e-02	2.11e-02	2.04e-02
		10.000.000	–	1.83e-01	1.80e-01	3.62e-02	2.09e-02	2.05e-02
	0.1	10.000.000	$\leq 1.98e-02$	1.04e-02	1.02e-02	2.15e-03	1.09e-03	1.07e-03
(2,1,[7])	0.3	10.000.000	–	4.98e-01	4.91e-01	1.18e-01	7.92e-02	7.74e-02
	0.2	10.000.000	–	8.37e-02	8.18e-02	1.80e-02	1.01e-02	9.90e-03
	0.1	10.000.000	$\leq 4.86e-03$	2.68e-03	2.62e-03	7.63e-04	3.80e-04	3.73e-04
(2,1,[10])	0.3	10.000.000	–	3.05e-01	2.98e-01	2.97e-02	1.85e-02	1.79e-02
	0.2	10.000.000	–	2.66e-02	2.57e-02	2.28e-03	1.21e-03	1.17e-03
	0.1	10.000.000	–	3.20e-04	3.09e-04	3.90e-05	1.80e-05	1.80e-05
(2,1,[14])	0.3	500.000	–	1.20e-01	1.16e-01	4.17e-03	2.54e-03	2.45e-03
	0.2	500.000	–	3.37e-03	3.24e-03	8.60e-05	4.20e-05	4.00e-05
	0.1	500.000	–	1.60e-05	1.60e-05	<1e-07	<1e-07	<1e-07
(2,1,[16])	0.3	500.000	–	1.56e-03	1.5e-03	2.00e-05	<1e-07	<1e-07
	0.2	500.000	–	<1e-07	1.5e-03	<1e-07	<1e-07	<1e-07
	0.1	500.000	–	<1e-07	<1e-07	<1e-07	<1e-07	<1e-07

Table 1. Replacing ROVA with soft input information and list decoding. Comparison of the results in [2] with hard decision decoding (HD) and soft decision decoding (SD), used with list size \mathcal{L} , where $\mathcal{L} = 1$ means no list decoding (all without ROVA). Extraction means the number of simulated key extractions.

Code	pr	Extractions	Ref.[2] $m = 1$	Ref.[2] $m = 3$	HD, $\mathcal{L} = 3$ $m = 3$	SD, $\mathcal{L} = 1$, $m = 3$	SD, $\mathcal{L} = 3$, $m = 2$	SD, $\mathcal{L} = 3$, $m = 3$
(2,1,[6])	0.3	500.000	4.11e-01	–	6.32e-01	7.81e-03	3.58e-02	4.93e-03
		10.000.000	–	–	6.32e-01	7.86e-03	3.59e-02	4.91e-03
	0.2	500.000	4.11e-01	–	1.84e-01	5.40e-05	1.35e-03	3.40e-05
		10.000.000	–	–	1.84e-01	6.10e-05	1.36e-03	3.20e-05
	0.1	10.000.000	$\leq 1.98e-02$	7.6e-07	3.27e-02	<1e-07	6.00e-06	<1e-07
(2,1,[7])	0.3	10.000.000	–	–	4.89e-01	2.19e-03	1.49e-02	1.32e-03
	0.2	10.000.000	–	–	1.05e-01	8.00e-06	3.44e-04	4.00e-06
	0.1	10.000.000	$\leq 4.86e-03$	<1e-07	2.66e-02	<1e-07	1.00e-06	<1e-07
(2,1,[10])	0.3	10.000.000	–	–	3.25e-01	4.70e-05	1.02e-03	2.85e-04
	0.2	10.000.000	–	–	6.43e-02	<1e-07	5.00e-06	<1e-07
	0.1	10.000.000	–	–	3.98e-02	<1e-07	<1e-07	<1e-07

Table 2. Adding the concept of using m readouts to the techniques used in Section 5.1.

Code	Extractions	SD, $\mathcal{L} = 1$	HD, $\mathcal{L} = 3$	SD, $\mathcal{L} = 3$	SD, $\mathcal{L} = 3$, $m = 3$
(3,1,[6])	10.000.000	3.98e-02	6.59e-01	2.24e-02	5.70e-05
(3,1,[7])	10.000.000	1.73e-02	5.93e-01	9.43e-03	6.00e-06
(3,1,[8])	10.000.000	9.72e-03	5.09e-01	5.14e-03	1.00e-06
(3,1,[9])	10.000.000	5.07e-03	4.28e-01	2.65e-03	<1e-07
(3,1,[10])	10.000.000	2.30e-03	3.39e-01	1.17e-03	<1e-07

Table 3. Using all response bits, even the unreliable ones.

Acknowledgement

The authors would like to thank George Yammine for support with the convolutional code library implemented by Walter Schnug, which serves as a basis for our own implementation.

References

1. M. Hiller, M. Weiner, L. Rodrigues Lima, M. Birkner, and G. Sigl, "Breaking Through Fixed PUF Block Limitations with Differential Sequence Coding and Convolutional Codes," in *TrustED*. ACM, 2013, pp. 43–54.
2. M. Hiller, A. G. Önalán, G. Sigl, and M. Bossert, "Online Reliability Testing for PUF Key Derivation," in *Proc. of the 6th Int. Workshop on Trustworthy Embedded Devices*. ACM, 2016, pp. 15–22.
3. G. E. Suh and S. Devadas, "Physical Unclonable Functions for Device Authentication and Secret Key Generation," in *Proceedings of 44th annual Design Automation Conference*. ACM, 2007, pp. 9–14.
4. J. Guajardo, S. S. Kumar, G.-J. Schrijen, and P. Tuyls, "FPGA Intrinsic PUFs and their Use for IP Protection," in *CHES*. Springer, 2007, pp. 63–80.
5. R. Maes, P. Tuyls, and I. Verbauwhede, "A Soft Decision Helper Data Algorithm for SRAM PUFs," in *IEEE Int. Symposium on Information Theory*. IEEE, 2009, pp. 2101–2105.
6. J.-P. Linnartz and P. Tuyls, "New Shielding Functions to Enhance Privacy and Prevent Misuse of Biometric Templates," in *Audio- and Video-based Biometric Person Auth.* Springer, 2003, pp. 393–402.
7. Y. Dodis, L. Reyzin, and A. Smith, "Fuzzy Extractors: How to Generate Strong Keys from Biometrics and other Noisy Data," in *Advances in Cryptology-Eurocrypt*. Springer, 2004, pp. 523–540.
8. S. Muelich and M. Bossert, "A New Error Correction Scheme for Physical Unclonable Functions," in *ITG SCC*, 2017.
9. P. Elias, "Coding for Noisy Channels," in *Proceedings of the Institute of Radio Engineers*, no. 3, 1955, pp. 37–46.
10. M. Hiller, L. R. Lima, and G. Sigl, "Seesaw: An Area-Optimized FPGA Viterbi Decoder for PUFs," in *17th Euromicro Conference on Digital System Design*. IEEE, 2014, pp. 387–393.
11. M. Bossert, *Channel Coding for Telecommunications*. John Wiley & Sons, 1999.
12. A. Viterbi, "Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm," *IEEE Transactions on Information Theory*, vol. 13, no. 2, pp. 260–269, 1967.
13. R. Maes, P. Tuyls, and I. Verbauwhede, "Low-Overhead Implementation of a Soft Decision Helper Data Algorithm for SRAM PUFs," in *CHES*. Springer, 2009, pp. 332–347.
14. J. C. Fricke and P. A. Hoeher, "Word Error Probability Estimation by Means of a Modified Viterbi Decoder," in *66th Vehicular Technology Conference*. IEEE, 2007, pp. 1113–1116.
15. G. Schmidt, V. R. Sidorenko, V. V. Zyablov, and M. Bossert, "Finding a List of Best Paths in a Trellis," in *IEEE Int. Symposium on Information Theory*, 2004.